

Metodologia internă privind cooperarea Centrului pentru Comunicare Strategică și Contracurarea a Dezinformării cu platformele tehnologice în procesul de sesizare a conținutului online

1. Dispoziții generale

1. Prezenta metodologie stabilește cadrul de cooperare dintre Centrul pentru Comunicare Strategică și Contracurarea a Dezinformării (în continuare – Centru) și platformele tehnologice, inclusiv Meta și Google/YouTube, în ceea ce privește sesizarea conținutului online, a conturilor, paginilor, canalelor sau a tiparelor de comportament digital care pot afecta subiecte de interes strategic aflate în mandatul Centrului.

2. Activitatea are caracter instituțional și se desfășoară exclusiv în raport cu domeniile și obiectivele prevăzute de cadrul normativ care guvernează activitatea Centrului. Cadrul legal relevant este dat de Legea nr. 242/2023 privind Centrul, astfel cum a fost modificată ulterior, precum și de Concepția de comunicare strategică și contracurarea a dezinformării, a acțiunilor de manipulare a informației și a ingerințelor străine. Totodată, având în vedere statutul Republicii Moldova de stat candidat la aderarea la Uniunea Europeană, Centrul poate utiliza, ca repere metodologice orientative, instrumentele relevante dezvoltate la nivelul EEAS și al Uniunii Europene privind FIMI, riscurile sistemice și integritatea proceselor democratice, fără ca acestea să producă efecte juridice directe în ordinea națională.

3. Cooperarea cu platformele nu reprezintă o activitate de moderare directă a conținutului și nu echivalează cu exercitarea unui control asupra deciziilor editoriale sau tehnice ale acestora. Rolul Centrului este unul de sesizare argumentată și documentată, în cazurile care intră în sfera sa de competență și care pot intra sub incidența regulilor comunității platformelor ori, după caz, a legislației în vigoare a Republicii Moldova. Meta și YouTube prevăd expres că aplicarea măsurilor asupra conținutului aparține platformei, potrivit propriilor politici și mecanisme de evaluare.

2. Temeiul de acțiune

4. Activitatea Centrului în raport cu platformele tehnologice are caracter de sesizare instituțională, desfășurată în baza mandatului legal al Centrului, a legislației în vigoare a Republicii Moldova și a regulilor comunității aplicabile fiecărei platforme.

5. Centrul nu recomandă și nu impune platformelor măsuri concrete asupra conținutului, ci transmite sesizări motivate cu privire la conținut, conturi sau comportamente care, în evaluarea sa, pot încălca regulile platformei, legile Republicii Moldova și pot afecta subiecte de interes strategic.

6. În toate cazurile, decizia de a elimina, restricționa, eticheta, suspenda sau de a nu interveni asupra conținutului aparține exclusiv platformei și moderatorilor acesteia, în conformitate cu politicile și procedurile proprii. Meta prevede că

standardele sale comunitare stabilesc ce este permis pe platformele sale, iar YouTube arată că materialele raportate sunt analizate conform propriilor Community Guidelines și că raportarea nu produce automat eliminarea conținutului.

3. Scopul activității de sesizare

7. Scopul activității de sesizare este de a permite Centrului să semnaleze platformelor cazuri relevante de dezinformare, manipulare informațională, conținut fabricat, impersonare, practici înșelătoare, comportament inautentic sau comportament inautentic coordonat, atunci când acestea pot prejudicia interesele naționale, pot afecta încrederea publică în instituțiile statului sau pot influența subiecte de interes strategic aflate în mandatul legal al Centrului.

8. Activitatea de sesizare vizează situațiile în care conținutul sau comportamentul online induce publicul în eroare în raport cu instituțiile publice, reprezentanții acestora ori cu subiecte de importanță strategică, inclusiv prin asociere falsă, atribuire mincinoasă de declarații, poziții sau acțiuni, conținut fabricat, manipulare tehnică, amplificare coordonată ori utilizarea unor rețele de conturi inautentice.

9. Activitatea de sesizare vizează, de asemenea, situațiile în care conținutul sau comportamentul online este susceptibil să afecteze procesele democratice, ordinea publică, stabilitatea socială ori economică, securitatea informațională, sectorul de securitate și apărare, parcursul european al Republicii Moldova, relațiile externe ori eforturile de consolidare a securității naționale în context regional.

10. Prejudicierea imaginii unei instituții publice nu constituie, în sine, temei autonom de sesizare. Sesizarea este justificată doar atunci când prejudiciul reputațional rezultă din falsificare, manipulare, impersonare, coordonare inautentică sau alte conduite care pot induce publicul în eroare și pot afecta un interes strategic aflat în mandatul Centrului.

4. Domeniul de aplicare

11. Centrul sesizează doar acele cazuri care se încadrează în subiectele de interes strategic prevăzute de lege și de documentele de politici relevante pentru activitatea sa. Activitatea de sesizare nu este generală și nelimitată, ci este restrânsă la conținutul și comportamentele care pot afecta, direct sau indirect, pilonii strategici protejați prin mandatul instituțional. Legea și Concepția aprobate la nivel parlamentar plasează activitatea Centrului în raport cu protejarea intereselor naționale, a proceselor democratice și a rezilienței în fața manipulării informaționale și a ingerințelor străine.

12. În sensul prezentei metodologii, subiecte de interes strategic reprezintă acele teme care, prin natura, amploarea sau potențialul lor de impact, pot afecta procesele democratice, funcționarea instituțiilor statului, ordinea publică, securitatea informațională, coeziunea socială, reziliența societală, securitatea și apărarea, parcursul european al Republicii Moldova, relațiile externe ori alte domenii aflate în sfera de competență a Centrului.

13. Prin urmare, Centrul nu sesizează orice conținut fals sau problematic identificat în mediul online, ci doar acele cazuri care au relevanță strategică și care au potențialul de a afecta procese democratice, încrederea în instituțiile statului, securitatea informațională, coeziunea socială, ordinea publică, reziliența societală sau alte domenii aflate în sfera sa de competență.

5. Principii de bază

14. Activitatea de sesizare se întemeiază pe principiile legalității, proporționalității, imparțialității, neutralității politice, documentării riguroase și delimitării clare între conținutul manipulator și exprimarea legitimă.

15. Centrul nu utilizează mecanismul de sesizare pentru a limita dezbateră publică autentică, pluralismul de opinii sau dreptul la critică. Nu constituie temei de sesizare simplul fapt că autorul conținutului este un partid politic, un actor electoral, o instituție media, o organizație a societății civile, un lider de opinie, un activist sau un critic al autorităților.

16. Centrul nu sesizează conținutul exclusiv pentru că acesta este ostil, critic, satiric, incomod, polemic sau nefavorabil unei instituții publice ori unui reprezentant al acesteia. Sesizarea poate interveni doar atunci când sunt prezente elemente de falsificare, inducere în eroare, impersonare, practici înșelătoare, coordonare inautentică sau alt comportament relevant în raport cu regulile platformei.

17. În aplicarea prezentei metodologii, Centrul va ține seama și de existența conținutului de interes public, a conținutului jurnalistic, documentar, analitic sau de dezbateră publică, inclusiv atunci când platforma prevede excepții sau evaluări contextuale pentru asemenea conținut. YouTube prevede expres excepții contextuale pentru anumite materiale cu valoare educațională, documentară, științifică, artistică sau de interes public.

18. Mecanismul de sesizare nu va fi folosit sub nicio formă în scop de intimidare, hărțuire, reducere la tăcere a unor voci legitime sau obținere a unor avantaje politice ori reputaționale. Meta prevede expres că sistemele sale nu pot fi folosite pentru a hărțui, intimida sau reduce la tăcere alte persoane.

6. Garanții instituționale de imparțialitate și control intern

19. Fiecare sesizare transmisă de Centru se bazează pe o fișă de evaluare standardizată, întocmită potrivit cerințelor și formularelor de sesizare aplicabile platformei vizate.

20. Înainte de transmitere, sesizarea este verificată intern sub aspectul relevanței pentru mandatul Centrului, al temeiniciei elementelor prezentate și al încadrării în regulile comunității platformei.

21. În cazurile care vizează actori politici, procese electorale, instituții media, subiecte de securitate sau alte dosare cu impact public ridicat, sesizarea este examinată cu un nivel sporit de rigoare, inclusiv prin analiza contextului, a autenticității materialelor și a delimitării dintre manipulare și exprimarea legitimă.

22. Nicio sesizare nu poate fi transmisă exclusiv în baza unui disconfort reputațional, a unei solicitări informale sau a apartenenței politice reale ori presupuse a autorului conținutului.

7. Obiectul sesizării

23. Pot face obiectul sesizării următoarele categorii de conținut sau comportament, atunci când ele intră în sfera de competență a Centrului:

a) conținut fals, manipulat sau fabricat care afectează încrederea publică în instituțiile statului ori induce în eroare opinia publică asupra unor subiecte strategice;

b) conținut care atribuie în mod fals declarații, decizii, documente, poziții sau acțiuni unor instituții publice ori reprezentanților acestora;

c) conturi, pagini, canale sau alte active digitale care se prezintă în mod fals drept instituții, reprezentanți oficiali ori surse legitime;

d) rețele de conturi sau pagini care acționează în mod coordonat și inautentic pentru amplificarea artificială a unor narațiuni manipulatorii;

e) comportamente digitale care pot fi încadrate, după caz, ca spam, practici înșelătoare, impersonare, fake engagement ori alte forme de comportament abuziv prevăzute de politicile platformei.

24. În cazul Meta, un reper important îl constituie politicile privind misinformation și inauthentic behavior. În cazul Google/YouTube, reperi relevante sunt politicile privind spam, deceptive practices and scams, impersonation, fake engagement și mecanismele de raportare a conținutului contrar politicilor interne sau legii aplicabile.

8. Limitele activității de sesizare

25. Centrul nu sesizează cazuri individuale sau personale care nu au relevanță strategică și nu afectează subiecte din mandatul său. De asemenea, Centrul nu intervine în litigii private, conflicte personale, dispute reputaționale individuale sau alte situații care nu produc efecte sistemice ori instituționale.

26. Totodată, Centrul nu sesizează conținutul care reprezintă doar o critică publică autentică, o opinie, o poziție politică, o dezbatere civică, un comentariu jurnalistic sau un material satiric, chiar dacă acesta este sever, incomod sau nefavorabil pentru imaginea unei instituții. Limita este depășită atunci când mesajul nu mai aparține sferei criticii legitime, ci intră în zona fabricării, falsificării, manipulării, atribuirii false, uzurpării de identitate ori amplificării coordonate prin rețele inautentice.

27. Centrul nu sesizează conținut veridic ori opinii critice legitime doar pentru că acestea afectează percepția publică asupra unei instituții. De asemenea, nu sesizează materiale jurnalistice, investigații, comentarii, opinii, satire sau dezbateri publice în lipsa unor elemente suplimentare de falsificare, manipulare tehnică, impersonare, practici înșelătoare ori coordonare inautentică. Meta și YouTube mențin politici distincte pentru autenticitate, înșelare și impersonare, nu pentru simpla critică ori dezbatere publică.

9. Rolul Centrului ca punct focal național

28. În calitate de punct focal național, Centrul poate intermedia solicitări de sesizare venite din partea instituțiilor statului din Republica Moldova, atunci când acestea privesc subiecte de interes strategic aflate în mandatul său și există indicii că regulile platformei ar putea fi încălcate.

29. Intermedierea de către Centru nu transformă sesizarea într-o instrucțiune obligatorie pentru platformă și nu creează o obligație de răspuns sau de acțiune din partea acesteia. Platforma decide în mod autonom dacă dă curs sesizării, dacă solicită informații suplimentare sau dacă nu intervine.

30. În exercitarea rolului de punct focal național, Centrul nu acordă tratament preferențial niciunei instituții, autorități sau categorii de actori. Intermedierea sesizărilor se realizează în baza aceluiași criterii de competență, relevanță strategică, standard de probă și compatibilitate cu regulile platformei.

31. Solicitarea unei instituții publice nu obligă Centrul să transmită automat mai departe sesizarea. În toate cazurile, Centrul efectuează propria evaluare independentă asupra încadrării cazului în prezenta metodologie.

10. Sesizările provenite de la organele de drept și alte autorități

32. Centrul poate fi sesizat de organele de drept sau de alte autorități competente în cazurile în care conținutul online identificat este susceptibil să încalce legislația în vigoare a Republicii Moldova și/sau intra sub incidența regulilor comunității unei platforme.

33. În asemenea situații, Centrul examinează și transmite mai departe sesizarea prin canalele aplicabile.

34. Această activitate nu substituie procedurile legale ale autorităților competente și nu reprezintă o calificare juridică definitivă a faptei sau a conținutului. Rolul Centrului rămâne unul de sesizare și intermediere în raport cu platforma.

11. Criterii de evaluare înainte de sesizare

35. Înainte de a transmite o sesizare, Centrul verifică dacă sunt întrunite următoarele condiții:

- a) cazul intră în mandatul legal și tematic al instituției;
- b) există indicii rezonabile și documentabile că acel conținut, conținut sau comportament poate încălca regulile comunității platformei;
- c) există un impact sau un risc de impact asupra unui domeniu strategic;
- d) sunt disponibile elemente minime de probă, precum linkuri, capturi, identificatori de conținut, exemple de conținut, tipare de publicare, corelații temporale, elemente de rețea sau alte indicii relevante;
- e) cazul depășește sfera criticii autentice și intră în zona falsificării, manipulării, impersonării, practicilor înșelătoare sau comportamentului inautentic coordonat.

36. În lipsa elementelor prevăzute la lit. b)–e), sesizarea nu se transmite.

37. În cazul suspiciunilor de comportament inautentic coordonat, analiza va urmări suplimentar existența unor indicatori precum sincronizarea publicării, repetitivitatea mesajelor, utilizarea unor identități false ori înșelătoare, amplificarea artificială a interacțiunilor, mascarea sursei reale a operațiunii, distribuirea coordonată pe mai multe conturi sau pagini și utilizarea unor active digitale create ori activate în scopul manipulării dezbaterii publice. Meta tratează distinct comportamentul inautentic, iar YouTube interzice fake engagement și diverse practici înșelătoare.

12. Fluxul procedural de sesizare

38. Procesul de sesizare include identificarea conținutului sau comportamentului relevant, verificarea relevanței sale pentru mandatul Centrului, încadrarea preliminară în raport cu regulile platformei, documentarea cazului, transmiterea sesizării și păstrarea evidenței asupra rezultatului.

39. Pentru fiecare sesizare transmisă, Centrul păstrează o evidență internă care include data, platforma, obiectul sesizării, temeiul de sesizare, autoritatea solicitantă, dacă este cazul, și răspunsul ori lipsa răspunsului din partea platformei. Evidența are rol de control intern, învățare instituțională și evaluare periodică a calității sesizărilor.

13. Transparență, raportare și revizuire internă

40. Centrul poate include, în raportările sale instituționale periodice, date agregate privind activitatea de sesizare, inclusiv numărul sesizărilor transmise, tipologiile principale, platformele vizate și rezultatele generale, fără a prejudicia activitatea operațională, protecția informațiilor sensibile sau alte obligații legale.

41. Centrul va efectua periodic o revizuire internă a practicii de sesizare, în scopul identificării eventualelor erori de încadrare, al uniformizării practicii și al prevenirii sesizărilor nefundamentate, disproporționate sau incompatibile cu principiile prezentei metodologii.

14. Exemple orientative

42. Exemple de cazuri susceptibile de sesizare:

a) Constituie caz susceptibil de sesizare distribuirea unui video fabricat, manipulat sau editat în mod înșelător, prezentat ca provenind de la o instituție publică ori de la un reprezentant al acesteia, prin care sunt comunicate în mod fals decizii, măsuri, declarații sau poziții inexistente.

b) Constituie caz susceptibil de sesizare distribuirea unor capturi de ecran false, documente fabricate, citații inventate sau alte materiale fals atribuite unei instituții publice ori unui reprezentant al acesteia, atunci când acestea sunt de natură să inducă publicul în eroare asupra unui subiect de interes strategic.

c) Constituie caz susceptibil de sesizare existența mai multor conturi, pagini sau canale aparent independente care publică simultan mesaje identice sau

substanțial similare, folosesc identități neverificabile ori false și urmăresc amplificarea artificială a unei narațiuni manipulatorii. Astfel de comportamente pot fi relevante mai ales din perspectiva politicilor privind comportamentul inautentic și coordonarea inautentică.

d) Constituie caz susceptibil de sesizare utilizarea unor conturi, pagini sau canale care se prezintă în mod fals drept instituții publice, reprezentanți oficiali, entități media legitime sau alte surse credibile, în scopul inducerii în eroare a publicului.

e) Constituie caz susceptibil de sesizare o campanie coordonată care, sub aparența criticii politice sau civice, utilizează conturi false, distribuire sincronizată, materiale fabricate și amplificare artificială pentru a crea percepția falsă că o instituție a statului a emis un act, a adoptat o decizie sau a formulat o poziție pe care, în realitate, nu a emis-o.

f) Constituie caz susceptibil de sesizare conținutul care vizează, în mod coordonat și manipulator, intimidarea sau hărțuirea unor instituții publice, a reprezentanților acestora ori a altor actori relevanți pentru procesele democratice, atunci când această conduită depășește critica legitimă și se manifestă prin amenințări, publicarea de date personale, atacuri repetate ori incitare la hărțuire. Meta și YouTube au politici distincte împotriva bullying-ului, hărțuirii, amenințărilor și doxxing-ului.

g) Constituie caz susceptibil de sesizare conținutul care instigă la ură, violență sau excludere împotriva unor persoane sau grupuri pe baza caracteristicilor protejate, atunci când acesta este utilizat în cadrul unor campanii care afectează coeziunea socială, ordinea publică sau alte subiecte de interes strategic. Meta interzice atacurile directe bazate pe caracteristici protejate, iar YouTube interzice promovarea violenței sau urii împotriva persoanelor ori grupurilor protejate.

h) Constituie caz susceptibil de sesizare conținutul care, prin tehnici de manipulare informațională, induce panică, neîncredere gravă sau reacții destabilizatoare în raport cu alegerile, securitatea publică, situațiile de criză, relațiile externe ori alte teme de importanță strategică.

i) Constituie caz susceptibil de sesizare utilizarea unor tehnici de spam, practici înșelătoare, scam, fake engagement sau alte metode artificiale de amplificare, atunci când acestea sunt folosite pentru a crește în mod neautentic vizibilitatea unor narațiuni manipulatorii. YouTube interzice explicit spamul, practicile înșelătoare și fake engagement.

j) Constituie caz susceptibil de sesizare conținutul care promovează acțiuni periculoase, violente sau ilegale, ori încurajează comportamente susceptibile să producă prejudicii grave, atunci când este folosit într-un context manipulator sau destabilizator. YouTube interzice conținutul care încurajează activități periculoase sau ilegale cu risc de vătămare gravă.

43. Exemple de cazuri care nu constituie, în sine, temei de sesizare

a) Nu constituie, în sine, obiect de sesizare un mesaj critic la adresa Guvernului, a Poliției, a unei autorități electorale sau a altei instituții, exprimat în termeni politici,

civici, academici sau jurnalistici, fără elemente de falsificare, manipulare, impersonare ori coordonare inautentică.

b) Nu constituie, în sine, caz susceptibil de sesizare o campanie politică, un comentariu jurnalistic, o investigație, o opinie, o dezbatere publică sau o critică severă, chiar sistematică, dacă aceasta nu implică materiale fabricate, identități false, practici înșelătoare sau alte încălcări ale regulilor platformei.

c) Nu constituie, în sine, temei de sesizare conținutul care reproduce afirmații controversate în scop informativ, analitic, documentar, educațional ori jurnalistic, în măsura în care contextul este clar și nu urmărește promovarea ori amplificarea manipulării. YouTube prevede explicit excepții contextuale pentru conținut educațional, documentar, științific, artistic și, în anumite cazuri, de interes public.

d) Nu constituie, în sine, caz susceptibil de sesizare un conflict strict individual, o dispută personală sau o problemă reputațională privată care nu are relevanță strategică și nu produce efecte sistemice ori instituționale.

15. Formulă de principiu

44. Centrul pentru Comunicare Strategică și Contracurare a Dezinformării sesizează platformele tehnologice exclusiv în privința conținutului, conturilor, paginilor, canalelor sau comportamentelor online care se încadrează în mandatul său legal, privesc subiecte de interes strategic și prezintă indicii rezonabile, documentabile și suficient motivate de încălcare a regulilor comunității aplicabile ori, după caz, de relevanță în raport cu legislația în vigoare a Republicii Moldova.

45. Activitatea Centrului are caracter de notificare și sesizare instituțională, nu de impunere a unei măsuri de moderare. Decizia de a acționa asupra conținutului raportat aparține exclusiv platformei, în conformitate cu propriile politici și proceduri.

46. Centrul nu sesizează critici publice autentice, opinii legitime, conținut de interes public, materiale jurnalistice ori cazuri strict individuale sau personale, ci doar acele situații care, prin caracterul lor falsificat, manipulator, impersonator, înșelător sau coordonat inautentic, pot induce publicul în eroare și pot afecta în mod semnificativ interesele strategice aflate în mandatul său.